Contemporary Educational Technology, 2025, **17**(4), ep606
ISSN: 1309-517X (Online)

OPEN ACCESS

Research Article

Enhanced predictive performance: A comparative analysis of ML and DL models using on augmented LMS interaction data

Kiran Fahd [1]
0009-0002-8213-3984

Shah J. Miah [1*]
0000-0002-3783-8769

[1] Newcastle Business School, College of Human and Social Futures, The University of Newcastle, Callaghan, NSW, AUSTRALIA
[*] Corresponding author: shah.miah@newcastle.edu.au

Citation: Fahd, K., & Miah, S. J. (2025). Enhanced predictive performance: A comparative analysis of ML and DL models using on augmented LMS interaction data. *Contemporary Educational Technology, 17*(4), ep606. https://doi.org/10.30935/cedtech/17453

ARTICLE INFO

Received: 27 Feb 2025

Accepted: 3 Jul 2025

ABSTRACT

The increasing reliance on the learning management system (LMS) in this era of digital education offers a vital source of student data that can be leveraged to predict student academic progress. Predicting student academic progress in higher education (HE) supports timely intervention and enhances student retention. This study develops and compares multiple machine learning (ML) and deep learning (DL) models to identify at-risk students based on students' interaction data with LMS by leveraging an integrated DSR methodology. Multiple predictive models are developed by incorporating data augmentation and balancing techniques to address class imbalance and enhance the accuracy of the predictive model. The study compares ten different models to achieve the highest classification accuracy in predicting students at risk of failing through the integration of through the integration of both ML and DL algorithms, including random forest, decision tree, convolutional neural networks, multi-layer perceptron, and long short-term memory (LSTM). The comparison results unscored the value of the DL based predictive model in the HE setting to precisely predict student academic performance, particularly the LTSM based model, which has the highest and nearly perfect accuracy. The existing LMS systems can incorporate this DL based predictive model to provide educational stakeholders with benefits and insights that support students' academic journeys and institutional success.

Keywords: machine learning, deep learning, design research, higher education, LMS data, student academic progress

INTRODUCTION

Learning management system (LMS) data provides a strong foundation for predicting student performance, a growing research area. LMS data is automatically generated through student behaviors and interactions with LMS that can be harnessed to anticipate their academic outcomes and implement appropriate support mechanisms if required. Early recognition of students at risk of failing allows for implementing appropriate intervention before students reach critical points of failure. Automated and data-driven approaches are increasingly being utilized in addition to traditional methods of identifying at-risk students, which often involve manual assessments, late detection, and delayed interventions (Tamada et al., 2021). Early and timely identification of at-risk students using data-driven information systems (IS) can support education providers in taking appropriate measures effectively to enhance student progress and prevent academic procrastination, thereby improving the overall effectiveness of the educational program.

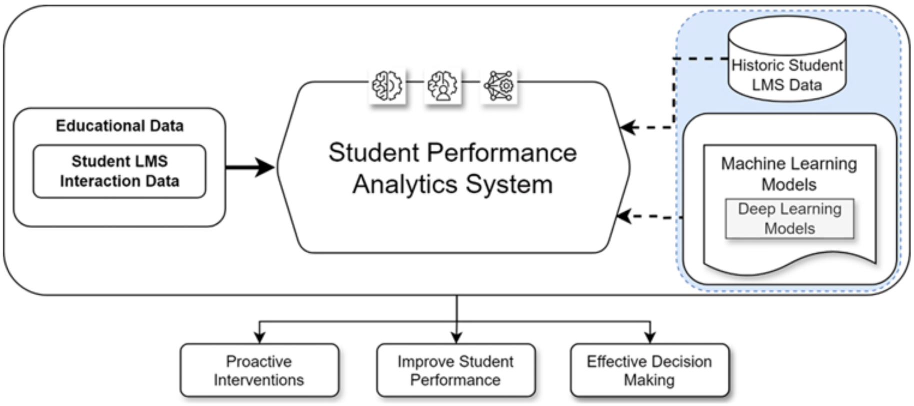

Figure 1. Overview of proposed educational technology (Source: Authors' own elaboration)

In recent literature, machine learning (ML) techniques have been used to analyze and detect patterns that can be transformed from educational data. Deep learning (DL) techniques, a subset of ML, employ neural networks (NN) for complex insight into datasets. Advanced data analysis techniques, such as ML and DL algorithms, are applied to LMS data to recognize patterns and make predictions based on the data. The application of ML has demonstrated potential in accurately predicting student performance and identifying struggling students requiring additional support (Masangu et al., 2020). Education providers can provide further support through quality learning and teaching to enhance students' academic performance. ML models utilize a wide range of attributes, including demographic information, educational records, and detailed logs of student interactions with LMS platforms to provide more nuanced insight into student behavior and performance (Badal & Sungkur, 2023; Maraza-Quispe et al., 2022; Pelima et al., 2024) and can uncover patterns and correlations that might not be evident by traditional manual data analysis. Furthermore, the continuous enhancement of these models, as research advances and accumulates more comprehensive datasets, promises even greater accuracy and practicality in the future (Saleem et al., 2021; Yağcı, 2022).

This study develops multiple ML and DL models utilizing a publicly available LMS dataset to precisely and accurately identify at-risk students. By following integrated DSR methodology, we developed an ML based model that uses LMS data and ML techniques to balance the dataset. Our ML model aims to enhance prediction accuracy, whether the student is at risk of failing or not, based on their interaction with the LMS. The main objective of this research is to achieve the highest classification accuracy of a ML based predictive model to identify struggling students by integrating existing class imbalance correction algorithms, ML algorithms and DL algorithms. Using the available data generated by LMS offers educators a practical and easily deployable tool. **Figure 1** provides an overview of our proposed educational technology to automatically detect at-risk students from LMS data so appropriate and timely measures can be applied to improve student learning. However, this paper focuses solely on building the model (highlighted in **Figure 1**), including developing a predictive model using LMS student interaction data and ML and DL algorithms.

The rest of the paper is arranged as follows: We first briefly explain the related work, standard DL and ML techniques, our dataset and performance metrics. Then, we describe the methodology of this study in depth. After that we present the discussion on the result of this study, and finally we summarize the research and suggest future directions.

STUDY BACKGROUND

Recent research has demonstrated the potential of ML and DL in identifying students who may be at risk of poor performance, thus enabling timely and effective educational interventions (Masangu et al., 2020).

Related Work

A few examples from existing studies about the application of ML in the educational sector, including but not limited to prediction of student enrolments, prediction of resources, career pathway recommendation application, adaptive tutoring, prediction of student academic progress or identification of at-risk students. **Table 1** provides a brief overview of related work in the education sector.

Table 1. Brief overview of related work in the education sector

Related work	Summary
A novel deep learning model for student performance prediction using engagement data (Fazil et al., 2024)	This study explores a DL approach to identify students needing additional support by predicting academic performance and utilizing engagement data from virtual learning environments.
A deep learning model to predict first to second year student retention (Beech & Yelamarthi, 2024)	This study investigates a neural network-based model that uses student academic data like GPA, enrollment data, and scores to predict first-year engineering student retention. It highlights the challenges in predicting student retention; the model accuracy range is 66.7%–95.2%.
A hybrid deep learning model to predict high-risk students in virtual learning environments (Syed Masood et al., 2024)	The study explored the DL models and CNN algorithms to forecast student performance and identify at-risk students, achieving an accuracy of 95.67% to provide early intervention strategies.
Video analytics in Moodle using xAPI (Judel et al., 2024)	This study explores the use of xAPI for Moodle video analytics. It logs, tracks and analyses detailed student video interactions, including played, paused, resumed, finished, and volume changes to analyze student engagement and predict academic performance. However, no AI or ML techniques are utilized in the study.
Predicting academic success of college students using machine learning techniques (Guanín Fajardo et al., 2024)	This study explores using different ML algorithms to predict academic success among college students. The study uses detailed student personal data, attendance and educational data to develop predictive models that can help identify students at risk of poor academic performance and prevent dropout.
A narrative review of students' performance factors for learning analytics models (Shafiq et al., 2023)	This study provides a comprehensive review of factors and attributes affecting student academic progress and highlights the role of the dataset generated by LMS and different ML based prediction models.
Improving models for student retention and graduation using Markov chains (Tedeschi et al., 2023)	The study explored the ML based learning assistant program and showed that the Markov model increases confidence and reduces biases for underrepresented students to improve the overall graduate rate.
Predicting student success using big data and machine learning algorithms (Ouatik et al., 2022)	The study explored complete student data (personal data, geographic data and academic evaluations) to develop and assess a ML based model to predict student academic performance. The support vector machine (SVM) base model outperforms other ML algorithms in the prediction rate.
Predicting student performance in a blended learning environment using learning management system interaction data (Fahd et al., 2021)	This study developed a ML based model by utilizing student LMS interaction data to identify students at risk of failing in a blended learning environment. The random forest based model demonstrated the highest accuracy among multiple other ML based models.

Our study aims to identify the most effective predictive model by comparing multiple ML and DL algorithms for the identification of at-risk students, contributing nuanced insights to the body of knowledge of the IS sector, especially in the context of blended or online learning, which incorporates various methods, including video-enhanced learning.

Existing literature highlights the effectiveness of various ML techniques, such as random forest (RF) or NN with LMS platforms dataset, in forecasting student outcomes. For example, ML models are trained and tested to analyze LMS clickstream data, which tracks every student interaction with the LMS platform. Different LMS users like educators, administrators, and students interact with the platform and generate This LMS data, generated by different LMS users like educators, administrators or students, includes user login frequency, student assignments submission, time spent or completion of various studies or activities, most visited study content, participation in discussions or extracurricular activities posts, all of which can be potentially indicative of a student's academic engagement and academic success or struggles (Alhothali et al., 2022; Aljaloud et al., 2022; Khan et al., 2023; Paramita & Tjahjono, 2021; Yakubu & Abubakar, 2022). Existing research commonly uses LMS clickstream data, including interactions such as login frequency and assignment submissions. However, our research focuses explicitly on time spent on LMS activities in general coupled with data related to video interactions, including detailed features, such as video plays, pauses, likes, segment views, and time spent on Moodle both on and off-campus, which provide a deeper insight into student engagement with video content, a vital component of learning using LMS. Analyzing video interaction features (play, pause, like, and segment), offers a more granular view of student behavior and uncovers nuanced patterns of student interaction with digital academic content growing trend in the education sector, potentially leading to more accurate predictions. Also, the primary focus of this study is leveraging attributes generated exclusively by

		PREDICTED VALUES		
		Positive	Negative	
ACTUAL VALUES	Positive	True Positive (TP)	False Negative (FN)	**Recall/Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP)	True Negative (TN)	
		Precision $\frac{TP}{(TP+FP)}$		**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

Figure 2. Performance metrics (Source: Authors' own elaboration)

LMS so that the implementation can be fully automated and seamlessly integrated into existing educational infrastructures to identify at-risk students.

Most existing studies have used supervised ML models to predict student academic progress. Decision tree-based methods are particularly favored due to their intuitive 'if-then' rule structures, which make them effective for predicting student progress and identifying at-risk students. Techniques such as decision trees, RF, J48, decision stump, OneR, NBTree, ID3, PART, naïve Bayes, NN, SVM, logistic regression, ZeroR, prism, multi-layer perceptron (MLP), and k-nearest neighbor have all been applied in this context. In this study, we have developed a predictive model using supervised ML algorithms and DL architectures to comprehensively compare their accuracy in predicting at-risk students. The ML and DL algorithms used in this study are given later.

In research studies, student final grades like GPA or final results have often been used as nominal attributes and predictors of student academic progress (Alfadhly, 2024; Aslam et al., 2021; Pelima et al., 2024). Therefore, this study has considered the final result of a student as the categorical variable for predicting student academic performance.

Data Augmentation Techniques

Datasets collected from educational technologies are often not correctly labelled and are thus imbalanced. According to existing literature, DL models do not consistently demonstrate reasonable prediction accuracy or even fail when applied to a small-scale dataset (Najafabadi et al., 2015). The synthetic minority oversampling technique (SMOTE) (Chawla et al., 2002) is a data augmentation technique for numeric data. SMOTE generates new synthesized data occurrences from the existing minority class instances instead of replicating data occurrences. It identifies the k-nearest neighbor of the minority class to create new data points along the line segment connecting randomly selected k-nearest neighbors of the minority class. However, SMOTE might add overlapping data points and noise as it does not account for the possibility that the neighboring data points can be from the majority class. This study applies data augmentation techniques to synthetically expand and balance the dataset.

Performance Metrics

The confusion matrix, as shown in **Figure 2**, a performance measurement for ML models, evaluates efficacy, effectiveness, utility, and usefulness of a predictive model. It breaks down the prediction of the model into true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) which provides calculations for several key metrics.

Classification accuracy is defined as ($\frac{TP+TN}{TP+TN+FP+FN}$) and demonstrates the effectiveness of the predictive model in a particular context. Classification accuracy is the ratio of correctly classified instances to the total instances. Other performance metrics are recall, sensitivity, and precision used in different contexts. Sensitivity or recall is defined as ($\frac{TP}{TP+FN}$) and gauges the ability of the model to identify positive predictions correctly, and a high recall value means few actual positives are missed. Precision is defined as ($\frac{TP}{TP+FP}$) and

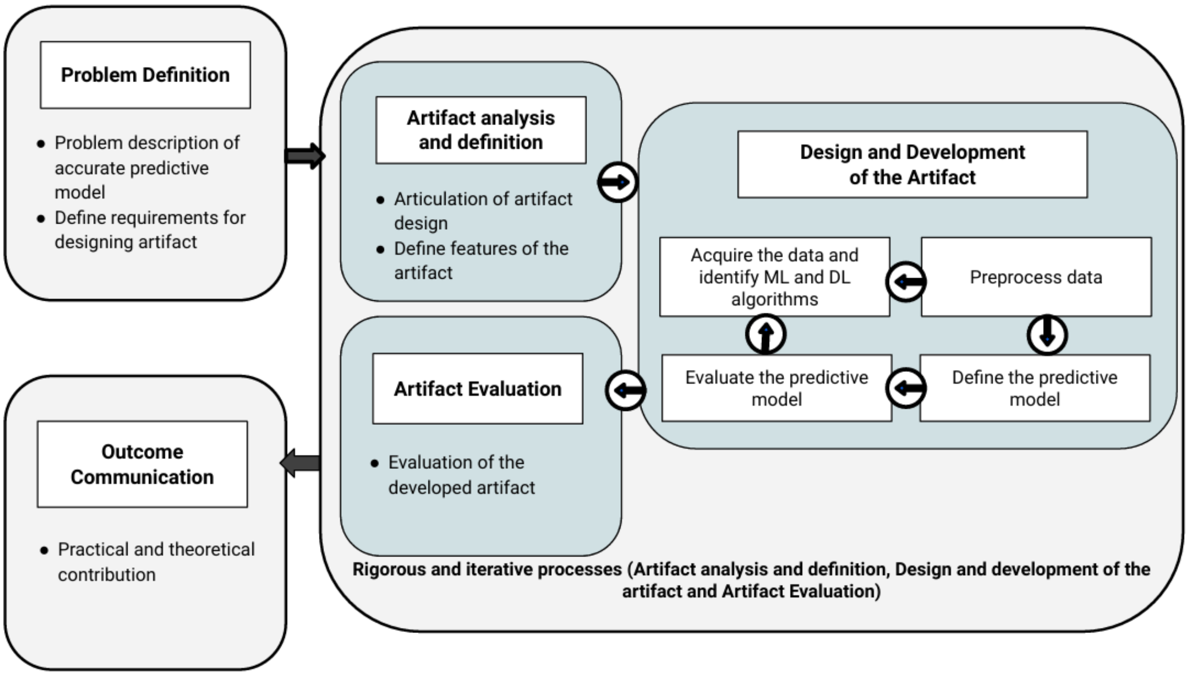

Figure 3. Integrated DSR methodology (Source: Authors' own elaboration)

assesses the reliability of positive predictions of the model, and high precision indicates that most positive predictions are truly positive. High accuracy, high precision and high recall are crucial for a well-performing model, as they show the capacity of the model to predict accurately and make reliable decisions. Precision and recall are compared in addition to classification accuracy in this study, as they are crucial for imbalanced datasets. An ideal model will return numerous results with high precision and high recall.

METHODOLOGY

Integrated DSR Methodology

The integrated DSR methodology is based on the similarities of DSR and DBR, which is followed to execute this study. The integrated DSR methodology consists of (as shown in **Figure 3**):

1. problem definition,
2. artifact analysis and definition,
3. design and development of the artifact,
4. artifact evaluation, and
5. outcome communication.

The study starts with a comprehensive problem identification and analysis of existing literature to identify the need for a highly accurate predictive model to predict students at risk of failing. The analysis drives the design requirements and objective of designing a predictive model using LMS interaction data from the literature. This analysis frames the design and development of the predictive model by incorporating ML and DL algorithms and data scaling and balancing techniques. Next, the study describes the application of these predictive models on the LMS dataset. It evaluates and compares the findings of the predictive model's performance using metrics like accuracy, F-measure, and precision, as well as comparing results before and after data scaling and balancing. In the final phases, the outcomes of the study are communicated as a contribution to the knowledge area to improve student academic outcomes.

Integrated DSR Artifact Development–Predictive Model

The ML lifecycle consists of four phases: data acquisition, data pre-processing, model training, testing, and model evaluation. A generic ML lifecycle is shown in **Figure 4**, which is leveraged in this study.

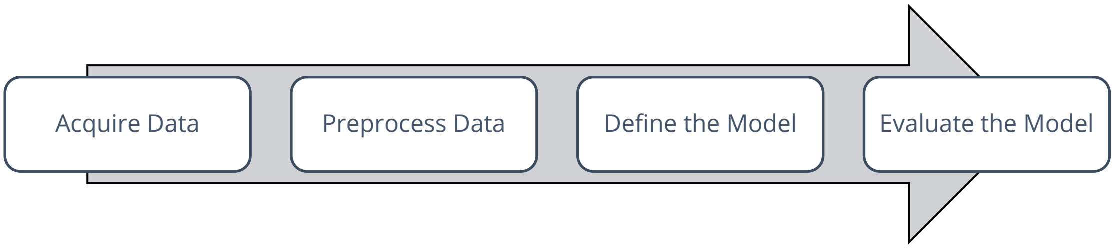

Figure 4. Generic ML life cycle (Source: Authors' own elaboration)

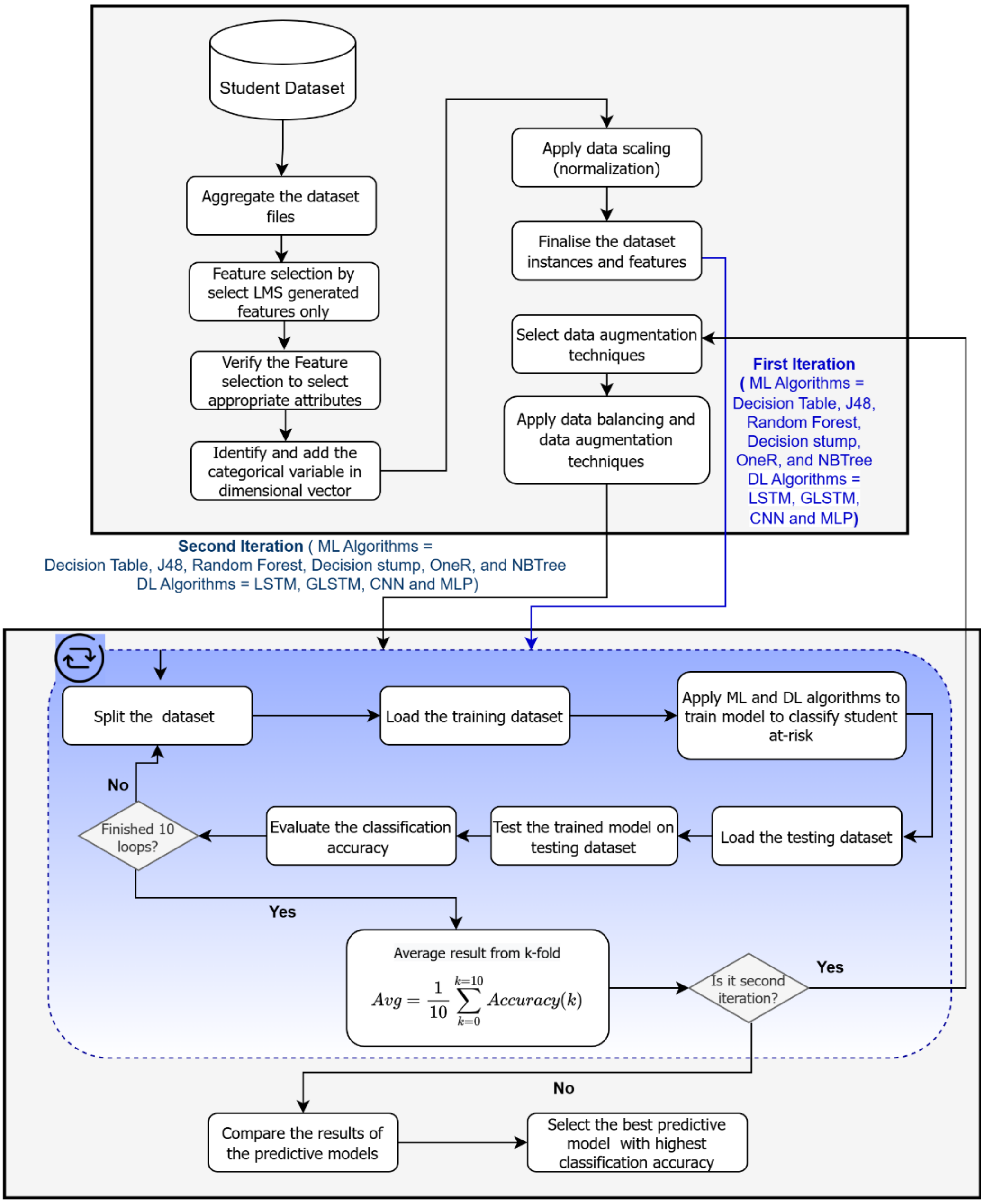

Figure 5. Integrated DSR artifact design and development of artifact– Predictive model (Source: Authors' own elaboration)

The detailed phases of training a predictive model to predict at-risk students using their interaction data with LMS, which will support educators in making proactive interventions and improving student academic performance, are given below. **Figure 5** depicts the step-by-step and complete process leveraging the ML lifecycle to demonstrate the methodology of this study.

Table 2. Brief description of dataset features

	Feature	Description	Statistics
1	Online_C	Time period in minutes of Moodle activities performed by students within the campus	Minimum: 2 Maximum: 597 Mean: 208.353 Standard deviation:124.018
2	Online_O	Time period in minutes of Moodle activities performed by students off-campus	Minimum: 0 Maximum: 587 Mean: 194.975 Standard deviation:131.131
3	Played	The count of the videos being played by the student	Minimum: 0 Maximum: 8 Mean: 2.132 Standard deviation: 1.761
4	Paused	The count of the videos being paused by the student	Minimum: 0 Maximum: 13 Mean: 2.184 Standard deviation: 2.604
5	Liked	The count of the videos being liked by the student	Minimum: 0 Maximum: 3 Mean: 1 Standard deviation: 0.967
6	Segment	The count of the specific portion of the video has been played by the student	Minimum: 0 Maximum: 7 Mean: 1.485 Standard deviation: 1.876
7	Result	Result of the students–Target feature & Pass (> 50 marks) & fail (< 50 marks)	0–100 & converted to pass (P) and fail (F)

Step 1. Acquire Data

The dataset used for this study is a publicly available dataset that is generated and collected from a real HE setting (Hasan et al., 2021). The dataset is publicly available under the creative commons attribution 4.0 international license (CC BY 4.0). It is the secondary dataset collected from students at a higher educational institute in Muscat, Oman, studying in a computing specialization from Spring 2017 to Spring 2021. No manual survey or interviews were conducted in the original data collection design study. According to Hasan et al. (2021), instructors obtained informed consent from students, coded student identities, and anonymized data using character generalization and marking methods. The dataset consists of 326 observations and 40 attributes divided into two categories: Student personal and academic information and student LMS activity and video interaction activity are collected from different platforms student IS, LMS and video interaction with eDify by students, e.g., the duration of activities performed in Moodle within the campus. For this study, we have only focused on LMS generated attributes to identify struggling students. **Table 2** briefly describes the seven attributes out of 40 attributes used in this study. The result attribute is defined as the target attribute. We acknowledge that this study is a secondary data analysis. Therefore, reflexivity, audit trail, and member checking steps are outside the scope of this study. Also, the data holds limited details of demographic and digital literacy information, which may introduce human biases and limit the generalizability of the model.

Step 2. Pre-Process Dataset

Dataset pre-processing involves data cleaning, feature selection, and scaling so that the data is clean, relevant, and in an appropriate format for training the model.

Feature selection

This study used the Python Matplotlib data visualization library to demonstrate the relative importance of each feature in a predictive model and verify our feature selection. As explained in the previous phase, this study solely utilized a numerical dataset generated by LMS. **Figure 6** depicts that all the selected seven features of the dataset have non-zero importance values, thus indicating that each feature contributes to

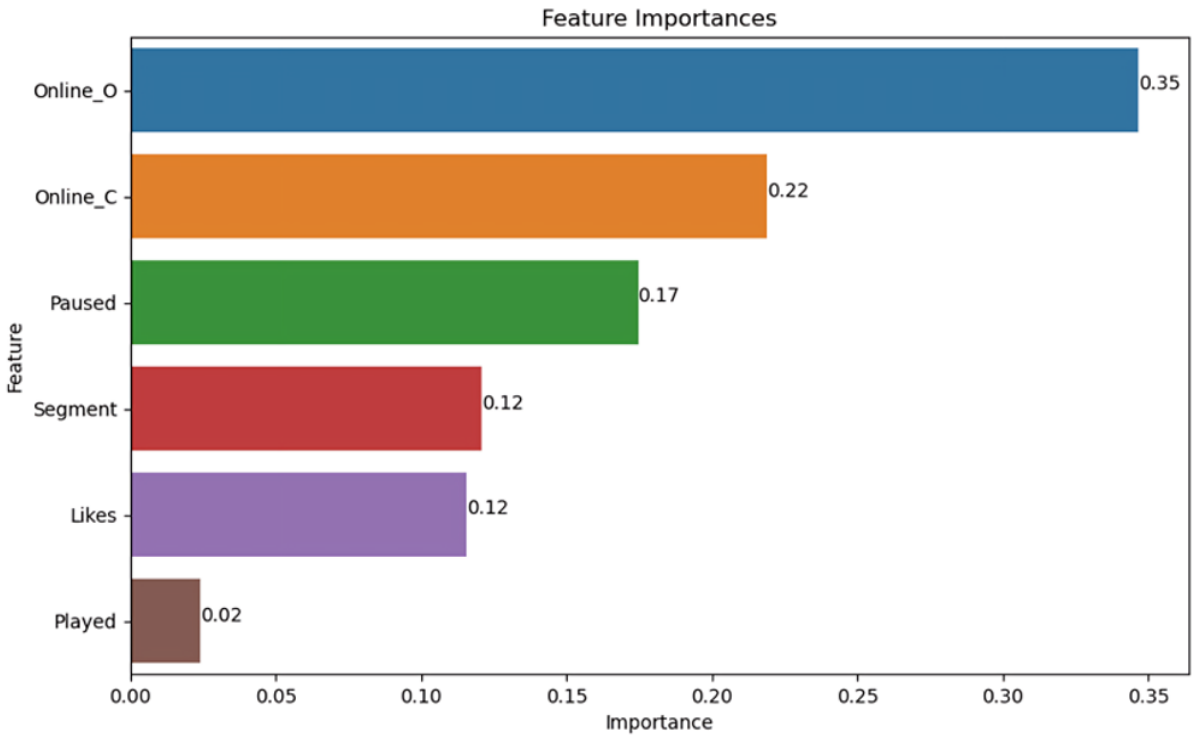

Figure 6. Feature importance plot (Source: Authors' own elaboration)

predicting the categorial variable, i.e., the result. Also, the distribution of important values across the features represents a well-balanced input feature set. Some features, such as online_O, have higher importance, and some features are moderately important, like paused. Other features, such as Played, still contribute, although to a lesser extent. This balanced distribution of the features can improve the robustness and generalization of the model.

Data scaling (normalization)

After the previous step, the feature set consists of one categorical variable with two classes (P and F) and six numerical features. The selected six features are not on a similar scale and consist of variable degrees of magnitude, range, and units. For example, the feature "played" represents quantity, i.e., a count of how many times the student played the video or units of another feature "online_C" are in minutes. Also, many predictive models perform better on scaled datasets. Two main data scaling techniques are often used, i.e., normalization (ranging between 0 and 1) and standardization (centered around the mean with a unit standard deviation–ranging between –1 and 1).

In this study, we have applied normalization to scale the dataset. Normalization ensures that all features are on the same scale and contribute proportionally to the training of the model. The MinMaxScaler class from the Scikit-learn Python library is employed to re-scale each feature of the dataset into a range of 0 and 1. The formula for normalization scaling is given below, where x is an individual value of each feature, and $\min(X)$ and $\max(X)$ denote the minimum and maximum values of that feature over the dataset.

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)}$$

The descriptive statistics of our dataset employing data scaling technique normalization is given in **Table 3**.

Data augmentation

Categorical variables of the dataset results and data distribution pass (P) or fail (F), 81% and 19% of the dataset, respectively. This ratio demonstrates that the data distribution is not the same for all classes, i.e., In these imbalanced datasets, the number of occurrences of the majority class (P) is higher than the minority class (F). This imbalance and biased dataset impact the performance of the predictive models. Often, the minority class (i.e., F) is ignored, which increases the likelihood of erroneous prediction of the minority class.

Table 3. Descriptive statistics of the dataset with seven features

Statistic	Online C	Online O	Played	Paused	Likes	Segment
Count	326.000000	326.000000	326.000000	326.000000	326.000000	326.000000
Mean	0.346811	0.332156	0.266488	0.168004	0.333333	0.212095
Standard deviation	0.208433	0.223392	0.220066	0.200271	0.322384	0.267956
Minimum	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25th percentile	0.190756	0.146508	0.125000	0.000000	0.000000	0.000000
50th percentile	0.312605	0.313458	0.125000	0.076923	0.333333	0.000000
75th percentile	0.487395	0.482112	0.375000	0.307692	0.583334	0.428571
Maximum	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

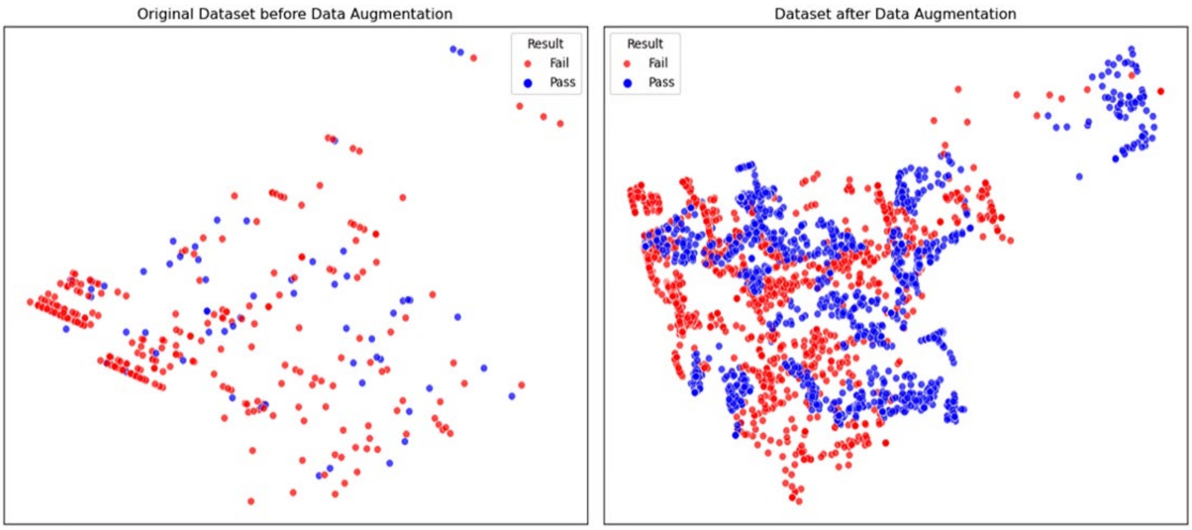

Figure 7. Datapoints before and after data augmentation (Source: Authors' own elaboration)

Table 4. ML and DL algorithms used to develop and evaluate the predictive model

	Algorithms used in this study
ML algorithms	Decision table, J48, RF, decision stump, OneR, and NBTree
DL algorithms	LSTM, GLSTM, CNN, and MLP

This study compares the prediction accuracy of the prediction of at-risk students with and without dealing with class imbalance by using the SMOTE data augmentation technique with ML and DL algorithms. **Figure 7** shows the data points before and after the data augmentation.

Define the Predictive Model

This step is an iterative process of applying the ML and DL algorithms on the dataset, training and testing predictive models before and after data balancing and augmentation. We selected supervised learning methods for ML algorithms to train and test the predictive model. We have used six tree-based classification methods, decision table, J48, RF, decision stump, OneR, and NBTree, to train and test the ML based predictive model. For DL algorithms, we have employed four DL algorithms, such as long short-term memory (LSTM), graves LSTM (GLSTM), convolutional neural network (CNN) and MLP, to compare and achieve the best predictive model based on performance metrics. **Table 4** summarizes the ML and DL algorithms used in this study to build and assess the predictive model.

In the first iteration, we defined the predictive models using ML and DL on the imbalanced dataset, as shown in **Figure 8**. In the second iteration, after rectifying the imbalanced dataset using data augmentation techniques, we redefined the models on the balanced and augmented dataset using the same ML and DL algorithms. Each predictive model is trained and tested three times to identify a more robust model. This study also evaluated the recall and precision measures of the predictive model with the highest accuracy.

Each predictive model based on ML and DL algorithms was trained and tested using k-fold cross-validation instead of the dataset train-test split to prevent model overfitting and to improve model generalization. In this

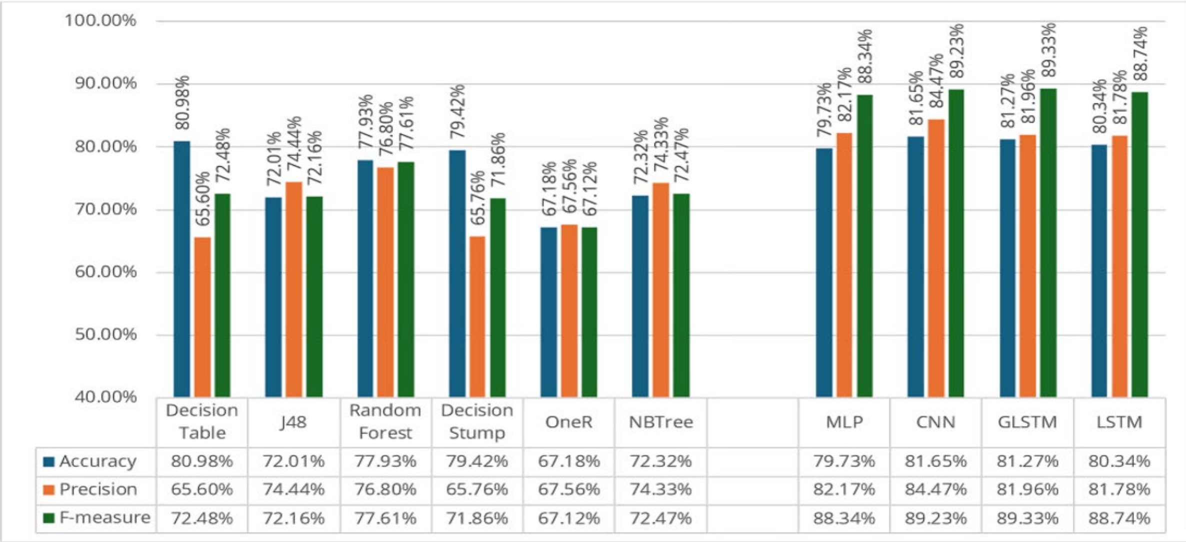

Figure 8. Predictive model performance metrics based on ML and DL algorithms without data augmentation techniques (Source: Authors' own elaboration)

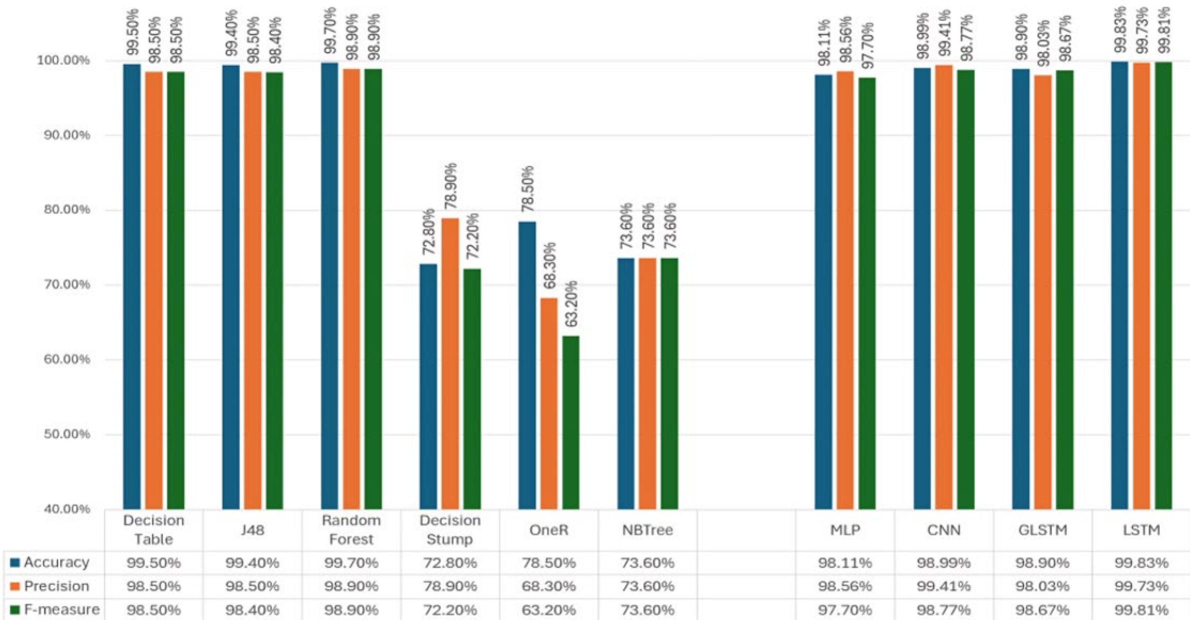

Figure 9. Performance metrics of the predictive model based on ML and DL algorithms and data augmentation techniques (Source: Authors' own elaboration)

case, the dataset is randomly split into k folds, and the model is trained on k–1 folds and tested on the remaining folds. In this study, the dataset is divided into 10 folds: k1 to k10, allowing each split to be used for training and testing. The model was trained in nine folds for each fold and tested on the 10th fold. After 10 iterations, the results were aggregated and averaged performance metrics from each fold. Different performance metrics, such as classification accuracy, recall, and precision, are compared for each ML and DL algorithm with and without augmented data. However, classification accuracy is only used to assess and select the best predictive model, as presented in **Figure 9**.

Table 5 compares performance metrics of ML and DL based predictive models before and after data augmentation techniques.

Table 5. Comparison of performance metrics before and after the application of data augmentation technique

Algorithm		Accuracy		Precision		F-measure	
		Before	After	Before	After	Before	After
ML algorithms	Decision table	80.98%	99.50%	65.60%	98.50%	72.48%	98.50%
	J48	72.01%	99.40%	74.44%	98.50%	72.16%	98.40%
	RF	77.93%	99.70%	76.80%	98.90%	77.61%	98.90%
	Decision atump	79.42%	72.80%	65.76%	78.90%	71.86%	72.20%
	OneR	67.18%	78.50%	67.56%	68.30%	67.12%	68.20%
	NBTree	72.32%	73.60%	74.33%	73.60%	72.47%	73.60%
DL algorithms	MLP	79.73%	98.11%	82.17%	98.56%	88.34%	97.70%
	CNN	81.65%	98.99%	84.47%	99.41%	89.23%	98.77%
	GLSTM	81.27%	98.90%	81.96%	98.03%	89.33%	98.67%
	LSTM	80.34%	99.83%	81.78%	99.73%	88.74%	99.81%

EVALUATION

The classification accuracy of our predictive model is vital to offering timely and appropriate support to struggling students. Accuracy for select ML and DL with and without data augmentation techniques are presented in **Figure 8**, **Figure 9**, and **Table 5**. Accuracy percentage represents the proportion of the correct classification and the total number of instances. Before the data augmentation technique, the decision table based model outperformed other ML based models and achieved the highest accuracy value of 79.4%, i.e., indicating that the decision table based predictive model is more accurate in predicting positive outcomes (identifying students at risk of failing). Similarly, CNN based model showed 81.65% accuracy from DL models. This accuracy is a significant performance metric of the predictive model, but it can be misleading if the dataset is imbalanced; therefore, precision and F-measure are also compared. Out of the ten predictive models, the precision (the ratio of TP to the sum of all positive instances identified by the predictive model) and F-measure are highest for RF and CNN based models. In our case, higher precision is preferable as it represents the low probability of FP, i.e., a low probability of at-risk students being identified incorrectly as not at-risk students. The value of the F-measure implies that predictive models based on RF and CNN algorithms have low FP and FN, i.e., low probability of not identifying students who are truly at-risk and higher accuracy of correctly identifying students at-risk. The decision table based model showed high accuracy, but its precision was relatively lower than that of the RF based model, indicating a higher rate of FP. In this case, the RF based model would be better for the predictive model as it has demonstrated balanced performance indicators, and the values of F-measure and precision are better than those of other algorithms, even though they showed slightly lower accuracy than the decision table based model. Overall, before the application of data augmentation techniques, the performance metrics of both ML and DL based models showed moderate effectiveness in their predictions. The accuracy of ML based models ranges from 67% to 80%, with similar trends for precision and F-measure metrics. However, complex DL based models demonstrated better and better performance across all metrics, i.e., accuracy, precision, and F-measure, ranging from 80% to 89%.

There is a substantial improvement across all the performance indicators of ML and DL based models after the application of data augmentation techniques (as explained earlier). The accuracy, precision, and F-measure for most predictive models, excluding decision stump, OneR, NBTree reached up to 99.8%, demonstrating the effectiveness of data augmentation in both ML and DL based predictive models in boosting robustness and generalization of a predictive model. Data augmentation offers more comprehensive data, improves class imbalance, reduces overfitting, improves generalization to unseen data and introduces data variability for better model learning. The predictive model based on a complex LSTM algorithm demonstrated remarkable consistency, maintaining high and balanced values across all performance metrics with approximately 100% (99.7%–99.83%) values. The stark contrast of performance metrics of ML and DL based predictive models before and after augmentation underscore the critical role of data augmentation in boosting the efficacy of both ML and DL predictive models, resulting in more accurate and reliable outcomes.

Table 6. Evaluation scenarios for predictive model evaluation

	Scenario	Data
1	Fail case × 2	0–49
2	Pass case × 2	50–100

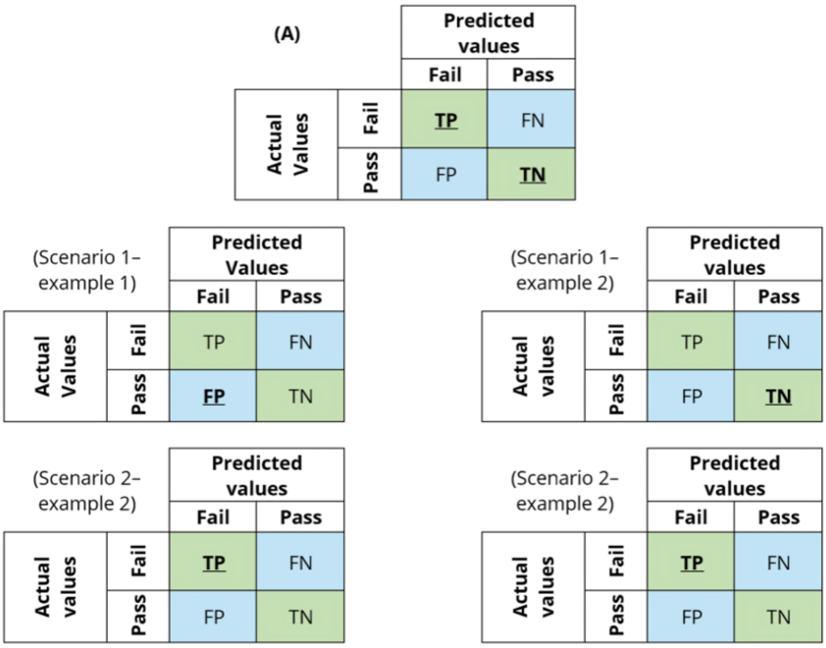

Figure 10. Evaluation of the predictive model using MFTs and validation dataset (Source: Authors' own elaboration)

In addition to the above, minimum functionality tests (MFT), which are also known as data unit tests, are used to evaluate the predictive model for specific cases found in the dataset. MFTs assess the effectiveness and efficacy of the select DL based predictive model for particular cases found in the dataset. The selected predictive model based on LSTM is evaluated and validated for specific instances from the dataset that meet the criteria in **Table 6**. In this study, we have extracted the validation dataset from the original dataset before training the model for the following cases that meet specific criteria as given in **Table 6**, and the scenarios are two different examples where to predict whether students would pass or fail a course based on interaction activities with LMS.

The validation dataset is obtained from the original dataset before training the model. The results of the predictive model evaluation using the MFT approach with two different scenarios for each class are given in **Figure 10**. A predictive model can be expressed into a 2 × 2 confusion matrix with the following four results:

1. True positive (TP green)–correctly identify at-risk students.
2. True negative (TN green)–correctly identify a student who is not at risk.
3. False positive (FP blue)–incorrectly identify a student as at-risk who is not at risk.
4. False negative (FN blue)–incorrectly identified a student not at risk who is at-risk.

The evaluation demonstrates that the predictive model accurately identifies the student at risk of failing in most scenarios. This evaluation outcome also answers the questions of the reliability of the data and the use of this model in the real world.

DISCUSSION

The outcome of the comparison of the predictive model's performance clearly demonstrates significant accuracy improvements after the application of data scaling techniques, i.e., the predictive accuracy increased dramatically, reaching as high as 99.83%. This outcome shows that data scaling and balancing techniques play a significant role in predicting students at risk of failing and strengthening the reliability and generalizability of the predictive model. Furthermore, the consistently high accuracy of the predictive models based on ten ML and DL algorithms suggests that this approach can be effectively applied in HE settings to provide accurate information to enable enhanced decision-making and interventions for academically struggling students.

The effectiveness of the predictive model is assessed by comparing its classification accuracy performance measure with existing state-of-the-art studies. To achieve this, we selected a few recent studies and presented their findings for comparison. These selected studies utilized different ML and DL algorithms to train and evaluate the models. Prior approaches such as a hybrid of decision tree (C4.5) stated 47% classification accuracy (Dang & Nguyen, 2022); a deep SVM with an improved CGAN accomplished 98.20% classification accuracy (Sarwat et al., 2022); an ensemble model of RF and SVM reached classification accuracy of 97.88% (Al-Ameri et al., 2024); an improved accuracy to 98.99% using an ensemble model of RF and SVM in another study (Saidani et al., 2024); and a standalone decision tree model reported a classification accuracy of 99.03% (Perkash et al., 2024). Based on a complex LSTM algorithm, our predictive model demonstrated superior classification accuracy compared to existing models, with an accuracy of 99.83%. These previous studies revealed that the advanced ensemble and DL based models perform better, but our approach of data augmentation with complex LSTM delivers additional performance gains.

This DL based predictive model embedded into educational technology is capable of automatically identifying students at risk of failing earlier in a program and supports educators in taking measures to improve the likelihood of student success. In this study, the dataset generated by LMS allows for completely automated educational technology that is cost-effective and can be easily customized across different institutions without manual human intervention. Also, integrating complex DL algorithms leads to more accurate and reliable identification of struggling students. It offers a proactive approach to educators to implement customized intervention approaches as part of positive strategies for students' ongoing success. The following are a few intervention strategies and suggestions:

1. Engage flagged students in a sequence of one-to-one consultations with study counsellors and mentors.
2. Offer literacy and numeracy support.
3. Extended academic sessions to guide identified students with learning content.
4. Offer tailored or individualized assessments while still attaining the learning objectives.
5. Monitor student participation and motivation with continuous reminders.

CONCLUSION AND FUTURE WORK

This study utilized integrated DSR methodology to compare ten different ML and DL models incorporating data scaling and balancing techniques to develop an educational system that timely and accurately identifies at-risk students. The comparison showed that performance indicators like the accuracy, precision, and F-measures of most DL and DL based predictive models, including decision stump, OneR, and NBTree reached up to approximately 99%, demonstrating that the data augmentation techniques enhance the robustness and generalization of the predictive model. The dataset utilized to train and test these models is comprised of data generated by LMS based on the activities and interactions of students with LMS. Appropriate intervention enables the implementation of remedial measures to reduce the probability of failure (and thus increase the retention rate). This study aims to support educators in predicting students at risk of failing to offer appropriate support programs and enhance academic performance. This timely identification of struggling students and support would improve student academic performance, leading to improved student retention and reduced student attrition, positively impacting student outcomes and reputation and the financials of educational institutions. This proactive intervention ultimately affects the economy of the nation by lowering the number of unpaid students so that students can repay their student loans.

In the future, more advanced data augmentation techniques, such as synthetic sequence generation and their impacts on the broad range of DL models, including LSTM and transformer variants, will be investigated and evaluated. Also, in the following phases, we will deploy an LSTM-based predictive model to assess its practical effectiveness in real-world educational settings. This model will empower an interactive dashboard designed specifically for educators and academic stakeholders that offers recommendations and plans based on risk assessment, such as early warning alerts, student mentoring sessions or customized learning resources and plans. To address real-world implementation challenges, we will explore a combination of engineering and methodological measures. For example, federated learning, differential privacy for data privacy and protection, and model compression and pruning to reduce inference latency facilitate smooth API integration with diverse LMS platforms. With these improvements, our predictive model will advance current DL research, in particular to deliver a technical, robust, and practical solution for educational stakeholders, for instance, to enhance supports of offering students' learning flexibility.

Author contributions: All authors were involved in the concept, design, data collection, interpretation, writing, and critically revising the article. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: The authors declared that ethics committee approval was not required for the study, as it is based on existing literature and a publicly available dataset.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Al-Ameri, A., Al-Shammari, W., Castiglione, A., NAPPI, M., PERO, C., & Umer, M. (2024). Student academic success prediction using learning management multimedia data with convoluted features and ensemble model. *Journal of Data and Information Quality*. https://doi.org/10.1145/3687268

Alfadhly, A. (2024). A comparative analysis for GPA prediction of undergraduate students using machine and deep learning. *International Journal of Information and Education Technology, 14*(2), 287–292. https://doi.org/10.18178/ijiet.2024.14.2.2050

Alhothali, A., Albsisi, M., Assalahi, H., & Aldosemani, T. (2022). Predicting student outcomes in online courses using machine learning techniques: A review. *Sustainability, 14*(10), Article 6199. https://doi.org/10.3390/su14106199

Aljaloud, A. S., Uliyan, D. M., Alkhalil, A., Elrhman, M. A., Alogali, A. F. M., Altameemi, Y. M., Altamimi, M., & Kwan, P. (2022). A deep learning model to predict student learning outcomes in LMS using CNN and LSTM. *IEEE Access, 10*, 85255–85265. https://doi.org/10.1109/ACCESS.2022.3196784

Aslam, N. M., Khan, I. U., Alamri, L. H., & Almuslim, R. S. (2021). An improved early student's academic performance prediction using deep learning. *International Journal of Emerging Technologies in Learning, 16*(12), 108–122. https://doi.org/10.3991/ijet.v16i12.20699

Badal, Y. T., & Sungkur, R. K. (2023, 2023/03/01). Predictive modelling and analytics of students' grades using machine learning algorithms. *Education and Information Technologies, 28*(3), 3027–3057. https://doi.org/10.1007/s10639-022-11299-8

Beech, M., & Yelamarthi, K. (2024). A deep learning model to predict first to second year student retention. In *Proceedings of the 2024 IEEE International Conference on Electro Information Technology* (pp. 324–327). IEEE. https://doi.org/10.1109/eIT60633.2024.10609845

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002, 06/01). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Dang, T., & Nguyen, H. (2022, 07/25). A hybrid approach using decision tree and multiple linear regression for predicting students' performance based on learning progress and behavior. *SN Computer Science, 3*. https://doi.org/10.1007/s42979-022-01251-5

Fahd, K., Miah, S. J., & Ahmed, K. (2021). Predicting student performance in a blended learning environment using learning management system interaction data. *Applied Computing and Informatics*. https://doi.org/10.1108/ACI-06-2021-0150

Fazil, M., Rísquez, A., & Halpin, C. (2024, 05/12). A novel deep learning model for student performance prediction using engagement data. *Journal of Learning Analytics, 11*(2), 23–41. https://doi.org/10.18608/jla.2024.7985

Guanín Fajardo, J., Guaña-Moya, J., & Casillas, J. (2024, 04/22). Predicting academic success of college students using machine learning techniques. *Data, 9*(4), 60. https://doi.org/10.3390/data9040060

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., & Sarker, K. U. (2021). Dataset of students' performance using student information system, Moodle and the mobile application "eDify". *Data, 6*(11), Article 110. https://doi.org/10.3390/data6110110

Judel, S., vom Felde, J., & Schroeder, U. (2024). Video analytics in Moodle using xAPI. *Technology, Knowledge and Learning, 29*, 1939–1963. https://doi.org/10.1007/s10758-023-09720-3

Khan, M., Naz, S., Khan, Y., Zafar, M., Khan, M., & Pau, G. (2023). Utilising machine learning models to predict student performance from LMS activity logs. *IEEE Access, 11*, 86953–86962. https://doi.org/10.1109/ACCESS.2023.3305276

Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., & Apaza-Huanca, J. M. (2022). Predictive model of student academic performance from LMS data based on learning analytics. In *Proceedings of the 13th International Conference on Education Technology and Computers*. https://doi.org/10.1145/3498765.3498768

Masangu, L., Jadhav, A., & Ajoodha, R. (2020). Predicting student academic performance using data mining techniques. *Advances in Science, Technology and Engineering Systems Journal, 6*, 153–163. https://doi.org/10.25046/aj060117

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1), Article 1. https://doi.org/10.1186/s40537-014-0007-7

Ouatik, F., Erritali, M., Ouatik, F., & Jourhmane, M. (2022). Predicting student success using big data and machine learning algorithms. *International Journal of Emerging Technologies in Learning, 17*, 236–251. ttps://doi.org/10.3991/ijet.v17i12.30259

Paramita, A. S., & Tjahjono, L. M. (2021, 09/20/). Implementing machine learning techniques for predicting student performance in an e-learning environment. *International Journal of Informatics and Information Systems, 4*(2). https://doi.org/10.47738/ijiis.v4i2.112

Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting university student graduation using academic performance and machine learning: A systematic literature review. *IEEE Access, 12*, 23451–23465. https://doi.org/10.1109/ACCESS.2024.3361479

Perkash, A., Shaheen, Q., Saleem, R., Rustam, F., Gracia, M., Alvarado, E., De la Torre Díez, I., & Ashraf, I. (2024, 04/29). Feature optimisation and machine learning for predicting students' academic performance in higher education institutions. *Education and Information Technologies, 29*, 21169–21193. https://doi.org/10.1007/s10639-024-12698-9

Saidani, O., Umer, M., Alshardan, A., Alturki, N., Nappi, M., & Ashraf, I. (2024, 03/18). Student academic success prediction in multimedia-supported virtual learning system using ensemble learning approach. *Multimedia Tools and Applications, 83*, 87553–87578. https://doi.org/10.1007/s11042-024-18669-z

Saleem, F., Ullah, Z., Fakieh, B., & Kateb, F. (2021). Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning. *Mathematics, 9*(17). https://doi.org/10.3390/math9172078

Sarwat, S., Ullah, N., Sadiq, S., Saleem, R., Umer, M., Eshmawi, A., Mohamed, A., & Ashraf, I. (2022). Predicting students' academic performance with conditional generative adversarial network and deep SVM. *Sensors, 22*, Article 4834. https://doi.org/10.3390/s22134834

Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2023). A narrative review of students' performance factors for learning analytics models. In *Proceedings of International Joint Conference on Advances in Computational Intelligence*. https://doi.org/10.1007/978-981-99-1435-7_23

Syed Masood, J. A. I., Chakravarthi, N., Asirvatham, D., Marjani, M., Shafiq, D., & Nidamanuri, S. (2024). A hybrid deep learning model to predict high-risk students in virtual learning environments. *IEEE Access, 12*, 103687–103703. https://doi.org/10.1109/ACCESS.2024.3434644

Tamada, M. M., Giusti, R., & Netto, J. F. D. M. (2021). Predicting student performance based on logs in Moodle LMS. In *Proceedings of the 2021 IEEE Frontiers in Education Conference* (pp. 1–8). IEEE. https://doi.org/10.1109/FIE49875.2021.9637274

Tedeschi, M. N., Hose, T. M., Mehlman, E. K., Franklin, S., & Wong, T. E. (2023). Improving models for student retention and graduation using Markov chains. *PLoS ONE, 18*(6), Article e0287775. https://doi.org/10.1371/journal.pone.0287775

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*(1), Article 11. https://doi.org/10.1186/s40561-022-00192-z

Yakubu, M. N., & Abubakar, A. M. (2022). Applying machine learning approach to predict students' performance in higher educational institutions. *Kybernetes, 51*(2), 916–934. https://doi.org/10.1108/K-12-2020-0865